DOCUMENT RESUME

ED 441 857 TM 030 892

AUTHOR	Adcock, Eugene P.; Phillips, Gary W.
TITLE	Accountability Evaluation of Magnet School Programs: A Value-Added Model Approach.
PUB DATE	2000-04-00
NOTE	24p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Ability; *Academic Achievement; *Accountability; Elementary Education; *Elementary School Students; Evaluation Methods; *Magnet Schools; Program Evaluation; Selection
IDENTIFIERS	Hierarchical Linear Modeling; *Prince Georges County Public Schools MD; *Value Added Model

ABSTRACT

School district accountability efforts have been challenged to develop procedures that determine the effects of their programs on student achievement within the typical implementation setting (i.e., nonrandom student assignment to nested school within program placements). Disentangling program effects from contextual and participant factors involves strategic considerations for data, design, and analysis. This paper presents an application of a two-level student-program hierarchical linear modeling (HLM) procedure to examine value-added effects of the Prince George's County, Maryland, Magnet School Programs at improving student academic performance. Examples of data handling and preparation for multilevel analysis are provided. Results from this study were primary considerations in the district superintendent's decisions regarding the expansion, continuation, and elimination of particular magnet school programs. The main findings in the study were: (1) overall, elementary students in magnet programs perform better than nonmagnet students; (2) this outcome is largely due to the fact that more able students are self-selected for the magnet program; and (3) when student ability is accounted for in the evaluation design, students in magnet school programs do not perform as well as students in nonmagnet programs. This last finding was further confirmed using unadjusted data in which talented and gifted (TAG) students in magnet programs did not perform as well as TAG students in nonmagnet schools. Appendixes contain tables of target data and program variables and a discussion of the statistical underpinnings of the HLM models used. (Contains 4 figures and 10 references.) (Author/SLD)

ACCOUNTABILITY EVALUATION OF MAGNET SCHOOL PROGRAMS: A VALUE–ADDED MODEL APPROACH

Eugene P Adcock
Prince George's County Public Schools

Gary W. Phillips[1]
National Center for Education Statistics
U.S. Department of Education

April, 2000

Abstract: School district accountability efforts have been challenged to develop procedures that determine the effects of their programs on student achievement within the typical implementation setting (i.e., non-random student assignment to nested school within program placements). Disentangling program effects from contextual and participant factors involve strategic considerations for data, design and analysis. This paper presents an application of a two-level student-program HLM procedure to examine the value-added effects of the Prince George's County Magnet School Programs at improving student academic performance. Examples of data handling and preparation for multilevel analysis methodology are provided. Results from this study were the primary consideration in the district superintendent's decisions regarding the expansion, continuation and elimination of particular Magnet school programs. The main findings in the study were 1) Overall, elementary students in magnet programs perform better than non-magnet students. 2) However, this outcome is largely due to the fact that more able students are self selected for the magnet program. 3) When student ability is accounted for in the evaluation design, magnet school programs do not perform as well as students in non-magnet programs. This finding was further confirmed using unadjusted data in which TAG students in magnet programs did not perform as well as TAG students in non-magnet schools.

Introduction

The 1999 Value–added Magnet school evaluation study is designed to assess the effectiveness of the Prince George's County Magnet School Programs at improving student academic performance. Student performance is measured by the six content areas of the MSPAP (reading, writing, language arts, math, science and social studies). To simplify the analysis a composite performance measure was derived from a linear transformation of the first principal component of the other six measures. This composite achievement score serves as the primary dependent outcome measure for this study. The evaluation in Phase I (fall, 1999) reported in this paper is limited to student cohort samples from the 1997-98 school year, and focuses on the MSPAP

[1] The discussion in this paper represents the views of the author and does not represent those of the U.S. Department of Education.

performance of third and fifth graders who were enrolled in their MSPAP administration schools for two consecutive years. The evaluation in Phase II (spring 2000) will extend the analysis of Magnet School Programs study to middle school with an analysis of students' 1998-99 school year grade eight MSPAP achievement scores.

This school-program evaluation shows how the value-added methodology entails data, design and analysis considerations for implementation. The data requirements which support value-added analysis are rigorous, but necessary for reliable and valid accountability evaluation. The quality and availability of data dictates the design process for controlling program treatment measures and sample cohort selection. The type of analysis model employed determines the level of precision and confidence yielded from the data analysis. This paper presents a discussion of the value-added approach, a data warehouse support system for multilevel evaluation analysis, and an application of a two-level student-program HLM procedure to examine the value-added effects of the Prince George's County Magnet School Programs at improving student academic performance.

VALUE-ADDED SCHOOL ASSESSMENT SYSTEM

Schools, districts and states have become increasingly interested in using student test results to assess the effectiveness of schools. Unfortunately, it is not widely appreciated that a proper school assessment system differs greatly from a simple aggregate of student test scores. It is important to draw a sharp distinction between a simple aggregate indicator based on test scores and indicators that extract measures of school effectiveness from test score data. That is, *school effectiveness* is the extracted measured contribution of *the school* from the student test results. The "value-added" model measures school effectiveness using a statistical model that accounts, to the extent possible, for all of the non-school factors that contribute to growth in student achievement. The objective is to statistically isolate the contribution of schools to student achievement growth from these other student and school contextual factors. A highly readable overview of value-added school assessment system is provided by Meyer (1997).

One simple way to think about value-added indices is that they are like unit pricing. The goal of unit pricing is to provide a standard value measure of a product by comparing the price of products per unit of weight, volume, or other amount. Comparing prices without and with unit pricing often leads to opposite results. For example, shampoo A as well as B may be good at reducing dandruff. If shampoo A were cheaper than shampoo B the casual shopper may decide that shampoo A is the better bargain. However, on closer inspection it is learned that shampoo B has considerably more volume of the shampoo than A. Controlling for the amount of shampoo, B becomes the better deal or value. The price comparison is made after standardizing or controlling the amount of the shampoo available. A shopper looking for value at the grocery store is implicitly, or explicitly using unit pricing.

The arena of educational accountability also has two types of shopping situations. The first type looks at schools with the highest average achievement test scores. These shoppers buy into the notion that schools with the highest test scores are the best "effectiveness" bargains. Evaluators of schools and program are the second type of shopper. Their job is to try to find out what makes some schools successful and others failures. They soon learn that high or low school performance is often more related to student demographics than to anything the school is doing (e.g., price is related to volume). School evaluators implicitly, or explicitly use the same methodology as unit pricing. They compare student achievement scores within the context of statistical controls for student demographic and school contextual factors. For example, since the population of students in a school is largely self selected, the effects schools on student achievement must be compared after standardizing or controlling for student ability. That is, *value added* is the contribution of *the school* after controlling for student demographic self selection factors.

Methodology

The purpose of the study is to provide an index of the effectiveness of Magnet School Programs after controlling for student demographic characteristics and school contextual variables. In addition, the study reports on school variables that contribute to the effectiveness of magnet schools, and suggest future data collections needed to better evaluate magnet schools.

In order to assess the unique value-added impact of the elementary school magnet programs on students' achievement, the 1998 magnet program MSPAP results are examined in the context of the following types of intervening variables: student characteristics such as students' first grade performance on OLSAT reasoning test and family poverty status; school context variables such as percent of student body receiving free and reduced meals; and school policy/practice variables such as teaching experience or college training level of the school's core teaching staff. These data supported the multilevel analysis model requirements used by this study to reveal the program impact on student achievement within the typical tangled context of school implementation.

Data

Data handling and data analysis are not distinctly different. Due to the increasing popularity of multilevel evaluation models (e.g., LISREL, AMOS, HLM) in school effects studies, the inherent perplexities of educational data are becoming more widely recognized (Bentler and Chou, 1988). On a practitioner level, school district evaluation offices need to get their data infrastructure in sync with the evaluation and accountability needs fulfilled by multilevel analysis models.

This study uses the Research, Evaluation and Accountability's relational data warehouse support system to provide two-level (student and school-program) data matrix files. The multilevel analysis files used analyzed with HLM are referred to as sufficient statistics matrices (SSM). The SSM file built for this study contains both level-1 summary statistics based on the constructed student data file which corresponds with each level-2 magnet school record, and the magnet school data. The resulting SSM file has as many records as the number of Level-2 magnet school program records .

Data Elements

The dependent outcome achievement variables for the Value-Added Magnet Program Evaluation study are the 1998 MSPAP results for grades three and five. The primary independent variable is the type of elementary school Magnet Program (i.e., non-magnet program, academic and traditional centers, communications and academic studies, creative and performing arts, french immersion, Montessori, science/math/tech, and talented-and-gifted/TAG). A dummy coding system was used to identify school's Magnet Program status for analysis. In addition to the primary independent and dependent variables, the SSM analysis files at Levels 1 and 2 are populated with student, program and school contextual variables which are considered important to providing deeper understanding of the process of schooling and determinants of magnet school program achievement. Table 1 shows descriptive statistics for the variables included in this study. The descriptive results are shown for the Level 2, or school level.

Table 1
School Level Descriptive Statistics For Program, Student and Teachers

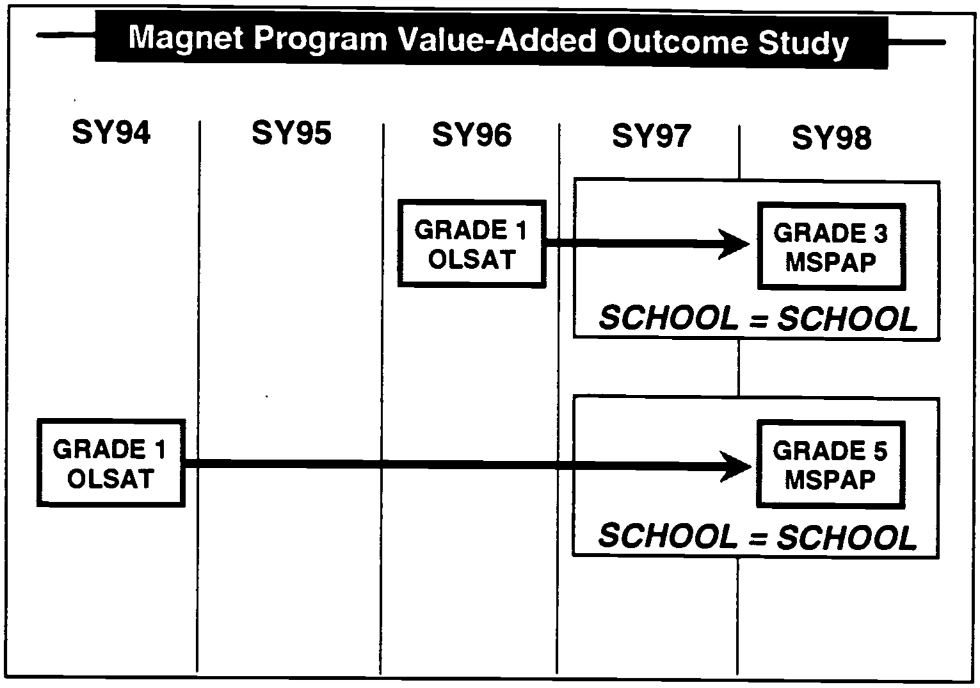
Dimension	Variables	N Schools	Minimum	Maximum	Mean	Std. Deviation
Pre/Post and VAI	Student ability Index (Pre-Test/Covariate)	117	88.8	123.0	100.63	7.25
	MSPAP Achievement (Posttest/Dependent)	117	470.4	556.2	511.12	18.99
	Value Added Index (HLM derived Effect Score)	117	-25.3	50.8	0.00	13.31
Students	Class Size	117	20.8	30.8	25.97	2.70
	Percent Afr-American	117	21.1	99.8	76.35	20.41
	Percent Poverty	117	9.2	85.6	49.53	19.71
	Percent ESOL	117	0.0	31.0	3.05	7.14
Teachers	Certification Index (1-5)	117	2.6	4.9	3.75	0.49
	Training Index (1-7)	117	1.3	3.7	2.05	0.37
	Years of System Service	117	2.1	21.6	9.56	4.04

Appendix 1 list the Level 1: Student, and Level 2: School-Program data elements extracted from data warehouse to create the SSMs analysis files. Operational definitions for the Level 1 and Level 2 variables in addition to coding schemes, and scaling parameters and conditions are also provided in the Appendix 1 table. This level of specificity in the development of the evaluation data set should meet standards of reproducibility and lend confidence to the information yielded by the proposed analysis.

Level 1: Student Sample

The Level 1 sample is composed of third and fifth grade students from the 1997-98 school year (SY98). Three selection criteria are used to select the Level 1 student cohort: posttest completion, pretest/covariate test completion, and at least two years of continuous program participation immediately prior to posttesting. Figure 1 presents a graphic representation of the sample selection criteria. The columns represent the school year (e.g., SY94 stands for school year 1993-94). The posttest condition, represented by the "Grade 3/5 MSPAP" boxes, requires sample students to have taken at least one of the six MSPAP content areas as a third or fifth grader in the spring of 1998. The pretest/covariate criteria, represented by the "Grade 1 OLSAT" boxes, requires sample students to have a first grade recorded school ability index (SAI) score on the OLSAT reasoning test. The OLSAT was administered in November 1993 for the SY98 fifth graders and November 1995 for the SY98 third graders. And finally, the length of treatment criteria, represented in Figure 1 by the "SCHOOL = SCHOOL" boxes, requires sample students to have belonged to their SY98 MSPAP administration school for two consecutive years immediately prior to posttesting (i.e., 1996-97 and 1997-98). Therefore, student level sample selection is dependent upon conditions of enrollment stability, first grade OLSAT test record, and MSPAP test scores for the student sample of third and fifth graders. For reasons that will be explained in the following Level 2 school-program sample section, all non-magnet students in non-dedicated magnet schools are excluded from the Level 1 sample, even if they meet the three selection criteria.

Figure 1



Level 2: Magnet School Program Sample

The Level 2 sample is composed of all 1997-98 elementary schools. Each elementary school is coded as a magnet school (N = 28) or as a non-magnet school (N = 89). The 28 magnet schools are further grouped by student program inclusion level, i.e., dedicated: all students in the school are magnet program participants; and, non-dedicated: only selected students within the school population are magnet program participants.

The non-magnet schools and the dedicated magnet schools are similar in that the entire school population participates in the schools' particular program orientation, i.e., all students in the school are either non-magnet or magnet. For analysis modeling purposes, this means that the variables used at Level 2 to characterize the school contextual factors that affect the models for student performance are based upon school-wide sample data (e.g., all students and all core teachers). The sampling situation for non-dedicated schools, however, is somewhat different. Because the evaluation study is trying to identify the effect of the magnet school program, independent of school contextual characteristics, it becomes necessary to separate the magnet program part of the non-dedicated school from the non-magnet part. Therefore, the non-dedicated school will be characterized entirely upon the Level 1 and Level 2 portion of the school which is part of the magnet program. Therefore, the non-magnet students will be excluded from the Level 1 sample in non-dedicated magnet schools. Only those teachers and school characteristics that can be specifically linked to the magnet student participants will be used to compose the Level 2 variables for the non-dedicated magnet schools. Table 2 presents an overall summary of the study purpose, sample, variables, design and analysis.

Table 2

ACCOUNTABILITY EVALUATION OF MAGNET SCHOOL PROGRAMS: A VALUE–ADDED MODEL APPROACH

PURPOSE: To estimate the extent to which Magnet Program experience contributes to student achievement (i.e., "value-added" analysis) independent of all other factors.

SAMPLE: Students – the 9,048 third and fifth grade MSPAP test takers who were enrolled in the same school for two consecutive years, and who have a first grade OLSAT-SAI score (i.e., from the 11/93 administration of the test for fifth graders, or the 11/95 administration of the test for third graders).

Schools – the 117 elementary schools that were in operation for both SY97 and SY98, and that could provide all pertinent student/school contextual data.

Note 1: For non-dedicated magnet schools (e.g., TAG and Montessori), student contextual variables are based on the magnet student population only.

Note 2: School contextual variables (e.g., class size and teacher college training) are defined at the school level rather than at the individual student or teacher level.

VARIABLES:

Dependent – the Spring 1998 scale scores for each of the six MSPAP content areas for both the third and fifth grades.

Covariate – each student's first grade OLSAT School Ability Index (SAI) score.

Independent – the school's magnet program status.

Contextual – student level variables such as race, poverty, and gender, and school level variables such as percent student poverty, class size, core teacher training and experience.

DESIGN:

Non-Randomized Comparison Group Pretest-Posttest Design

	Pretest		Treatment		Posttest
Experimental Group (Mag)	O_1	=>	X	=>	O_2
Comparison Group (Non-Mag)	O_1	=>	C	=>	O_2

O_1 – the OLSAT (nonverbal) results from the 11/94 or the 11/96 administration of test for first graders

X – two years of magnet program treatment from the Fall of 1996 through the Spring of 1998

C – two years of non-magnet (i.e., comprehensive) program treatment from the Fall of 1996 through the Spring of 1998

O_2 – the MSPAP composite results from the six content areas administered to third and fifth graders during May 1998

ANALYSIS Value-added analysis using 2-level HLM analysis to measure the impact of a magnet school program on student MSPAP achievement independent of student ability and other student and school contextual variables. The results will be expressed in terms of each schools' Value-Added Index (VAI), which shows the impact of a school's academic program on its MSPAP scale score.

Analysis

Hierarchical linear modeling, or HLM, has recently replaced multiple linear regression (MLR) as the method of choice for school-effects research (Mendro, Webster, Bembry & Orsak, 1995; Meyer, 1997). HLM estimates linear equations that explain outcomes for students within schools as a function of the characteristics of the school as well as the characteristics of the students. Because HLM involves the prediction of achievement of students who are nested within schools which in turn may be nested in larger groups (e.g., programs, clusters, and districts), it is ideally suited for use in education.

Bryk, Raudenbush, Seltzer, and Congdon (1988) cited four advantages for the appropriateness of HLM analysis for school-effects research. First, HLM can explain student achievement and growth as a function of school-level characteristics while taking into account the variance of student achievement within schools. Second, it can model the effects of student background variables such as student family poverty status and student mobility on test performance within schools and explain differences in these effects between schools using school characteristics (e.g., teacher competence, and percent of school on free/reduced meals). Third, HLM can model the between-and within-school variance simultaneously and thus produce better estimates of student achievement outcomes. Finally, it can produce better estimates of the predictors of student outcomes within schools by using information about these relationships gained from other schools within the district (i.e., provide a "level playing field").

The analysis for this study uses a two-level hierarchical linear model (HLM) to provide a value-added index of program effectiveness. At the student level of the model the analysis will explore the effects of aptitude (as measured by the student's 1st grade OLSAT school ability index), gender, minority status, and poverty (as measured by participation in free and reduced lunch). At level 2 (the school-program level) the analysis will review the effects of program attributes, teacher characteristics (teacher training, mobility, and experience) and other school contextual variables (school poverty, demographic characteristics). Appendix 2 of this paper provides an overview of the basic statistical analysis modeling procedure that will be employed in this magnet school program evaluation study.

The two-level model used in this study examines differences in achievement among elementary students nested within magnet school programs. Considering the magnet school program's complex student selection process (i.e., lottery, choice, recruitment, and sometimes testing) and two-level implementation procedures (i.e., dedicated and non-dedicated), HLM is the analysis procedure of choice to accommodate the nested program operations structure and nonrandom sample situation. That is, after modeling the magnet school program in terms of student and school variables, HLM provides a means of predicting how variables at a higher level can affect the models for student performance at a lower level. For the interested reader, a number of reference works are available on the application of HLM for providing deeper understanding of the processes of schooling and determinants of school achievement presented in Adcock and Phillips, 1997; Arnold, 1992; Bryk and Raudenbush, 1992; Phillips and Adcock, 1996 and 1997.

Results

There were three general findings from the study. These were:

1. Most of the elementary school students in Magnet Programs in Prince George's County Public Schools performed as well, or better than non-Magnet students on the MSPAP.

2. However, this outcome is largely due to the fact that more able students are in the Magnet School Program.

3. When student ability is accounted for in the evaluation design, students in the Magnet School Programs do not perform as well as students in the non-Magnet programs. This same finding was further confirmed when comparisons were made between TAG students in the Magnet School Program and TAG students not in the non-Magnet program.

Each of these findings will now be discussed in more detail.

<u>Finding 1</u>

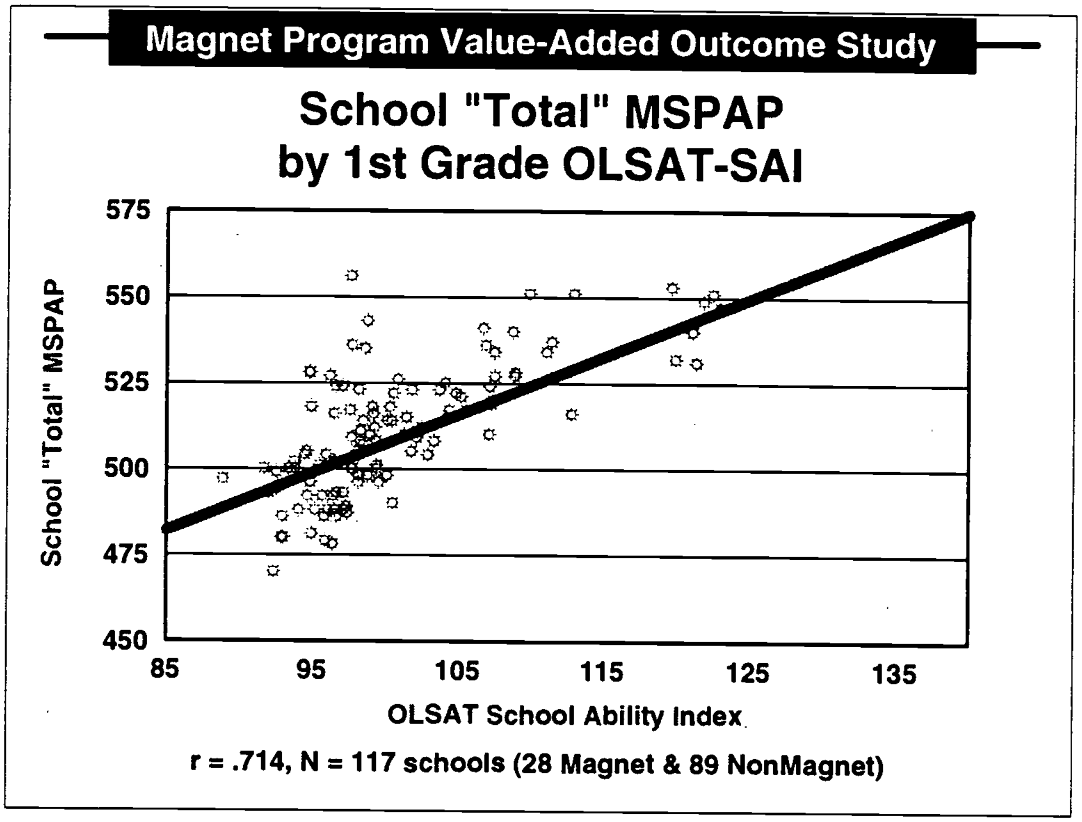
The first finding was that there is no significant difference in overall performance in the elementary school students in Magnet Programs and non-Magnet programs on the MSPAP (Mean Magnet = 516.8, Mean non-Magnet = 509.3, F=3.4, $p<.069$). This finding is based on a simple comparison of mean MSPAP scores between magnet-and non-Magnet students with no controls for the student's prior ability.

<u>Finding 2</u>

Figure 2 shows the relationship between the school-programs' overall achievement on the MSPAP composite[2] in each school (the first principal component of the 6 subject areas) and the mean IQ of the students in the program (as measured by the OLSAT). As can be seen, there is a strong correlation ($r = .714$) between the achievement of the students in the program and the prior ability of the students in the program. The highest achieving schools have the most students with high prior ability. Note, however, that the prior ability of the students has nothing to do with program effectiveness or the value-added by the program. Almost half (49%) of the variance in program achievement can be explained by the prior ability of the students. A reasonable conclusion from Figure 2 is that we will never be able to measure the value-added by the Magnet Programs until we can disentangle program effectiveness from the confounding influence of the students prior ability.

[2] The "overall MSPAP composite" score is derived from the first principal component factor of the 6 MSPAP subject areas. For interpretive purposes, the factor score was re-scaled to the observed MSPAP mean and standard deviation.

Figure 2

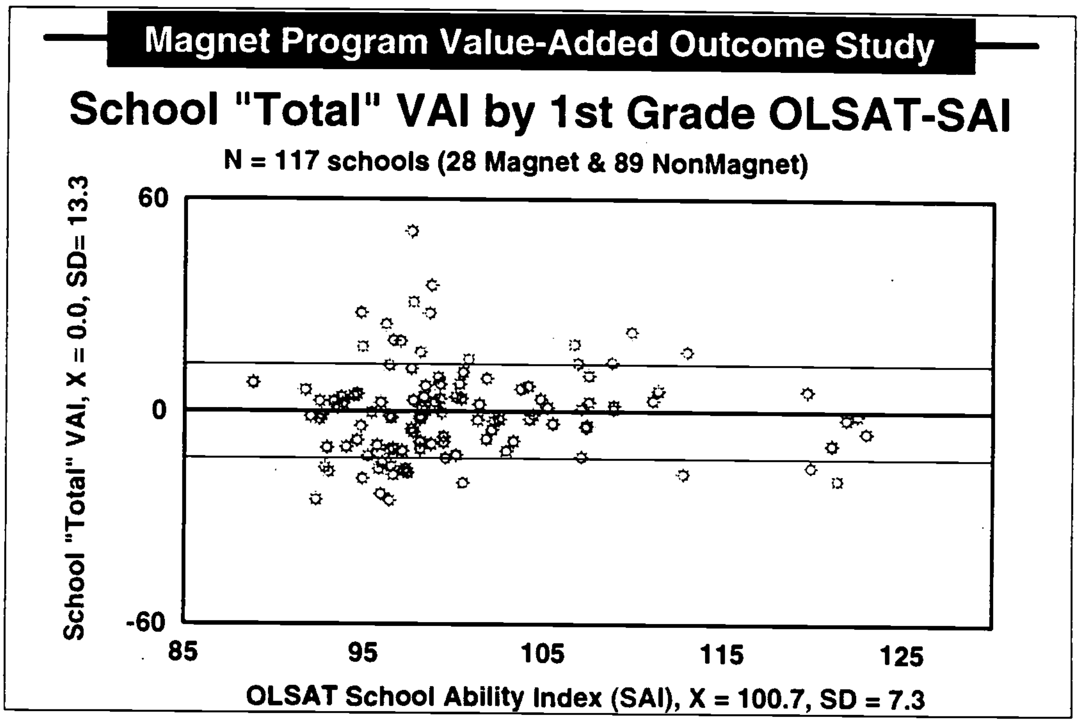


Finding 3

The third finding was obtained by fitting a two-level hierarchical linear model (HLM) to the data (see equation 5 -7 in the Appendix 2). In this model statistical procedures were used to control for each student's prior ability, and the average prior ability of the group of students in the program. The final result of the HLM analysis was to create a Value-Added Index that shows the effect of the Magnet Program on the program's students after controlling for the prior ability of the students in the program. It should be noted that the Value-Added Index is in the same metric as the original MSPAP. The Value-Added Index may be thought of as that portion of the MSPAP score that is due to the value-added by the Magnet Program.

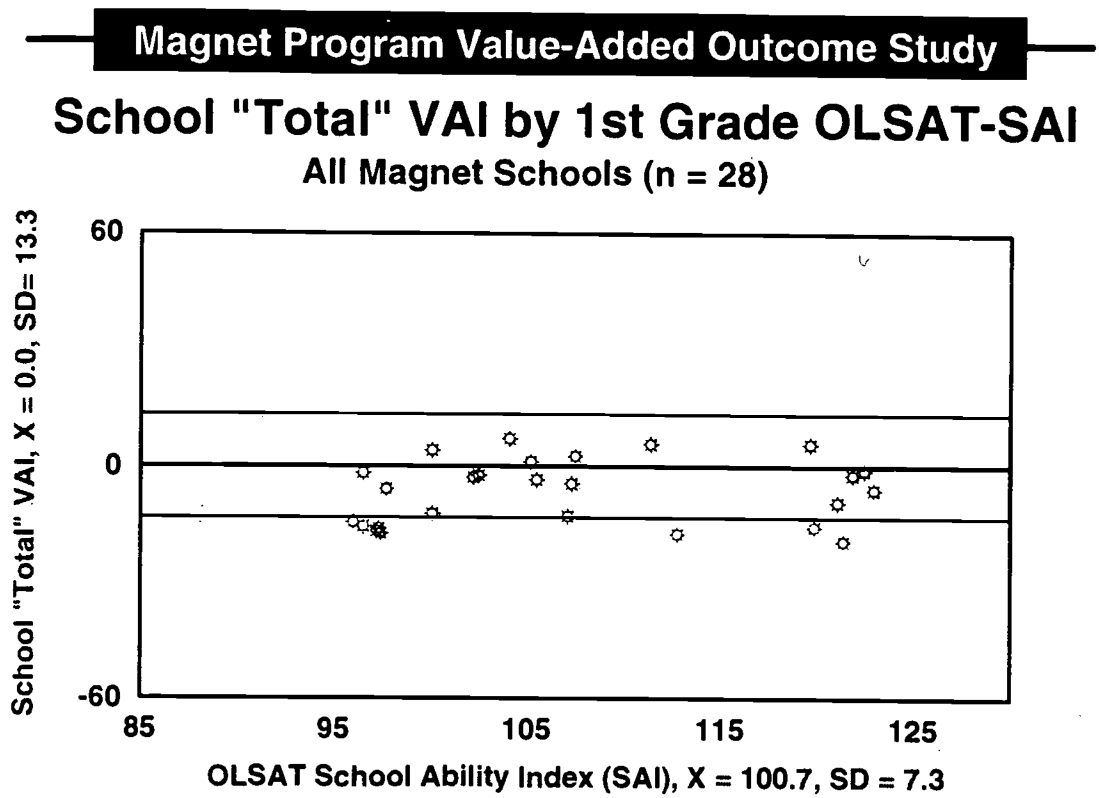
Figure 3 displays the relationship between the programs Value-Added Index and the mean IQ of the students in the program (as measured by the OLSAT). As can be seen, there is no correlation between the Value-Added Index and prior ability of the students in the program (the effects of prior ability has been controlled). Programs that have a Value-Added Index more than one standard deviation above the county mean are the most effective schools, and programs which have Value-Added Indices below one standard deviation are the least effective. Programs within one standard deviation are essentially comparable to the county average. The important thing about the Value-Added Index is that it is a measure of the success of the program that no longer depends on the ability of the students. Some programs are successful even though they have the lowest ability students. Others are not successful, even though they have the highest ability students. The Value-Added Index removes the artificial advantage enjoyed by the programs with the most able students, and removes the excuse used by programs with the least able students.

Figure 3



After controlling for prior ability, the average score of students in the Magnet program is significantly lower than the non-magnet schools (Mean Value-Added for Magnet program = -6.4, Mean Value-Added for non-Magnet student = 2.0, F = 9.1, p<.003). Figure 4 displays the 28 magnet school programs, extracted from the overall 177 school results in Figure 3. This graphic shows that the magnet school programs' results are either comparable to the county average or, in some cases, below the county average in regards to the Value-Added Indexes.

Figure 4



The HLM procedure used in finding 3 "statistically" controls for the prior ability of the students. A good validation of this controlling procedure is available in the evaluation study for the TAG Magnet School Programs. To conduct this evaluation we compared the MSPAP achievement scores for students in the TAG Magnet School Program against TAG students in the non-Magnet Schools. In this analysis there is no need to "statistically" control for prior ability because the all TAG students (Magnet and non-Magnet) have comparable high IQs.

Summary and Future Research

School district accountability efforts have been challenged to develop procedures that determine the effects of their programs on student achievement within the typical implementation setting (i.e., non-random student assignment to nested school within program placements). Disentangling program effects from contextual and participant factors involve strategic considerations for data, design and analysis. A two-level student-program HLM procedure used to measure the value-added effects of the Prince George's county Magnet School Programs at improving student academic achievement yielded the following findings:

1) Overall, elementary students in magnet programs perform better than non-magnet students.

2) However, this outcome is largely due to the fact that more able students are self selected for the magnet program.

3) When student ability is accounted for in the evaluation design, magnet school programs do not perform as well as students in non-magnet programs. This finding was further confirmed using unadjusted data in which TAG students in magnet programs did not perform as well as TAG students in non-magnet schools.

Results from this study were reported in the district's Magnet Program Evaluation Task Force's report: "Summary of Findings from the Magnet Report Card." (Fall, 1998). The Value-Added analysis results played a major role in the Task Force's recommendations. The initial recommendations identified three programs for "Replication," two programs for "Improvement/Enhancement," five schools from various Magnet Programs for "Immediate Intervention," and two programs for "Replace with New Theme Or Eliminate." Based upon these recommendations, the Superintendent set one year achievement and program implementation accountability targets which selected school-programs must meet in order to continue.

The evaluation reported in this paper is limited to elementary school student cohort samples from the 1997-98 school year, and focuses on the MSPAP performance of third and fifth graders who were enrolled in their MSPAP administration schools for two consecutive years. The next phase of this evaluation (spring 2000) will extend the outcome study of Magnet School Programs to the middle schools with an analysis of students' 1998-99 school year grade eight MSPAP achievement scores.

References

Adcock, E. P. & Phillips (1997). *Measuring School Effects With Hierarchical Linear Modeling: data Handling and Modeling Issues.* Paper presented at American Educational Research Association, 1997 Annual Meeting, Chicago, *Session # 16.2, Hierarchical Linear model Applications,* SIG/Multiple Linear Regression: The General Model.

Arnold, C. L., (1995) *Using HLM and NAEP Data to Explore School Correlates of 1990 Mathematics and Geometry Achievement in Grades 4, 8, and 12: Methodology and Results*, National Center For Education Statistics (NAEP), Research and Development Report, Jan., 1995.

Bryk, A.S., & Raudenbush, S.W., *Hierarchical Linear Models: Application and Data. Analysis Methods,* Sage, 1992.

Bryk, A. S., & Weisberg, H. I. Value-added analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics,* 1976, 1, 127-155.

Davidson, F. (1996). Principles Of Statistical Data Handling, Sage Publications, Inc., Thousand Oaks, CA.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28,* 465-498.

Haseltine, R., Winkler, L., & Adcock, E. P. (1998) *Getting Ready for Hierarchical Linear Modeling: Data Prep 101.* Paper presented at the American Educational Research Association AERA Annual Conference, San Diego, CA.

Phillips, G. W., & Adcock, E. P., (1997). *Practical Applications of Hierarchical Linear Models to District Evaluations,* Journal of Multi-Linear Regression Viewpoints (Spring, 1997).

Phillips, G. W., & Adcock, E. P., *Using Hierarchical Linear Models to Evaluate Schools*, Paper presented at American Educational Research Association, 1996 Annual Meeting, NY, Presentation 45.13, *Business Meeting/HLM Discussion/Symposium*, "Multiple Linear Regression: The General Model" (4/11/96).

Raudenbush, S. W., & Willms, J.D. *The estimation of School Effects,* Journal of Educational and Behavioral Statistics, Winter 1995, Vol. 20, No. 4, pp. 307-335.

Appendix 1

Target Data for Magnet School Evaluation Study (Phase 1)
List of Student (Level 1) and School/Program (Level 2) Variables

Variable	*Definition*	*Comment*
Unit of Analysis Variables		
Student	A **student** is a third or fifth grader during the 1997-98 school year, took the MSPAP during the 1998 school year, took the OLSAT as a first grader (5^{th}:11/93, 3^{rd}:11/95), belonged to his/her MSPAP administration school for three consecutive prior years, and matriculated within a designated Magnet Program type (or Non-Magnet Program = control sample).	"Student" is a *Research, Evaluation and Assimilation Database* (READ) warehouse system data element containing <u>all</u> of the following: a valid student ID number, a first and last name, a valid race code, a valid gender code, and a date of birth. For evaluation purposes, **student** is often conditionally defined in terms of the study in question (i.e., in terms of a year cohort, the student's teacher, school, program, course, test, etc.).
Teacher	A **core teacher** is a school-based, certified or provisional certified classroom teacher employed on the last day of the school year and who has the assigned responsibility to provide ***students*** instruction and assign course grades in one or more of the core academic subject areas of language (including reading and English), mathematics, science, or social studies.	**Core teacher** is a READ system data element. Depending on the nature of an evaluation study, the number of core teachers for an entity (e.g., school, program, or year cohort) can be either observed or derived from other accounting sources.
School	For the purposes of this study, school refers to an elementary level school that provides all of its students with magnet program curriculum (dedicated), some of its students with magnet program curriculum (non-dedicated), or none of its students with magnet program curriculum (non-magnet)	School is a READ system data element represented by a valid system identification code, have one or more teachers to give instruction, are located in one or more buildings, and have an assigned administrator.

Variable	*Definition*	*Comment*
Magnet School Type (including non-magnet)	Elementary schools that offer some, all, or none of their students the special or unique subjects, activities and/or learning opportunities for the following magnet program curriculum: Non-Magnet, Academic-Traditional/Classical Centers, Communications & Academic Centers, Creative & Performing Arts, French Immersion, Montessori, Science/Math/Tech, and Talented & Gifted (TAG).	Magnet schools are specialized programs in selected elementary, middle, and high schools. The PGCPS essential curriculum is included in all magnet programs, along with a range of special or unique subjects, activities and/or learning opportunities.
	Level I Student Variables	
Gender	Student gender status is coded as follows: 0 = Female, and 1 = Male.	
Minority	Student minority status is coded as follows: 0 = Non-African American, and 1 = African American	
Poverty Status	Student eligibility for federally subsidized meal program based on family economic status. Student family poverty status is coded as follows; 0 = Not eligible for Free/Reduced meal program, or 1 = eligible for F/R meal program.	
School Ability Index (SAI)	Students' SAI score on the Otis-Lennon School Ability Test (OLSAT)	The students are administered the OLSAT during November of first grade. Thus, the third and fifth grade cohorts were administered the OLSAT in 11/96 and 11/94, respectively.
Mathematics MSPAP Scale Score	Students' SY98 MSPAP scale score in mathematics content area	MSPAP results collected for SY98 third and fifth graders included in the Magnet Program Study cohort samples, only.
Reading MSPAP Scale Score	Students' SY98 MSPAP scale score in reading content area	(See above)

Variable	*Definition*	*Comment*
Science MSPAP Scale Score	Students' SY98 MSPAP scale score in science content area	(See above)
Language MSPAP Scale Score	Students' SY98 MSPAP scale score in language content area	(See above)
Writing MSPAP Scale Score	Students' SY98 MSPAP scale score in writing content area	(See above)
Social Studies MSPAP Scale Score	Students' SY98 MSPAP scale score in social studies content area	(See above)
Level II Magnet School-Program Variables		
School Level Teacher College Training	The **school level teacher college training score** is the mean of the academic training index of the core teachers *belonging* to the school site on the last day of the school year. Each core teacher is assigned an academic training index score on a seven point scale: 1=Bachelors, 2=Bachelors+30 course credit hours (cch), 3=Masters/Equivalent, 4=Masters+15(cch), 5=Masters+30(cch), 6=Masters+60(cch), 7=Doctorate.	Data source is Personnel Office college training code for individual teachers.
School Level Teacher Certification Status	The **school level teacher certification status** is the proportion of the core teaching *belonging* to the school site on the last day of the school year that have Standard Professional-I instructional certification, or higher.	The school teacher certification status is derived as the mean of dichotomously coded teacher status: Provisional = 0, and Standard (or higher) = 1.

Variable	*Definition*	*Comment*
School Level Teacher Service Years School	The **school level teacher service years index** is the mean of the total number of years that the core teacher *belonging* to the school site on the last day of the school year was employed at the "event" school. Years of experience operationally defined as "belonged" to that school based upon their most recent occurrence of continuous employment in that school.	Only the last continuous "belonging" period in the "event" school is considered. Data source is summarized Personnel Office records. An "event" school is location of interest to the evaluation study.
School Level Teacher Service Years in System	The **teacher service years in the system** is calculated as the average total number of years that core teachers belonging to the event school have been employed as teachers (provisional or certified) **in the system** based upon their most recent occurrence of continuous employment in the system.	Data source is Personnel Office for individual teachers. Only the last continuous belonging period is considered
School Level Minority Status (African-American)	The percentage of the total SY98 MSPAP examinee sample in third and fifth grade cohort who are African-American for each elementary school (e.g., minority value of 45% represents 45% of the target sample is African-American, and 55% are non:African-American.	School aggregate means of Minority = 1 and Other = 0 are actually proportion values of study students who are African-Americans.
School Level Class Size [See special note section below]	The **school level class size** is a constructed class size statistic derived from the total number of students enrolled on the last day of the school year divided by the school teacher count.	"School Teacher Count" is a READ-defined data element that represents the number of "classroom teachers" at a school site. Because "School Teacher Count" is used in the denominator, this constructed class size for each school is a robust estimate of each school's average instructional unit, thus an improvement over the student-teacher ratio typically used in the literature (Ferguson, 1991).
School Level Mathematics MSPAP Scale Score	The school's average unweighted third and fifth grade student performance for MSPAP mathematics	A few elementary or "combination" schools did not have both third and fifth grade levels.

Variable	*Definition*	*Comment*
School Level Reading MSPAP Scale Score	The school's average unweighted third and fifth grade student performance for MSPAP reading	(See above)
School Level Science MSPAP Scale Score	The school's average unweighted third and fifth grade student performance for MSPAP Science	(See above)
School Level Language MSPAP Scale Score	The school's average unweighted third and fifth grade student performance for MSPAP language	(See above)
School Level Writing MSPAP Scale Score	The school's average unweighted third and fifth grade student performance for MSPAP writing	(See above)
School Level Social Studies MSPAP Scale Score	The school's average unweighted third and fifth grade student performance for MSPAP social studies	(See above)
School Level Teacher Service Experience	The **school level teacher years of experience** is the recorded sum years of teaching credit assigned for teaching (provisional and certified) regardless of location. Does not have to be continuous years of teaching employment.	Source is summarized Personnel Office data.

Appendix 2

Value-Added Magnet School Evaluation Study

This appendix is intended to provide the basic statistical underpinnings of hierarchical linear models (HLM) that was employed in this evaluation of magnet school programs. A simple version of the HLM was used in the study.

Two-level Example of the HLM

To facilitate understanding we will "illustrate" all points with a simple 2-level HLM model with only one independent variable at both the student and school levels. In the actual study reported in this paper 7 dependent variables were used with 7 independent variables. We will also adopt the widely used notation provided by Bryk and Raudenbush (1992) with the addition that random variables will be underlined.

Level-1

$$\underline{Y}_{ij} = \underline{\beta}_{0j} + \underline{\beta}_{1j}(\text{StudentAbility}_{ij}) + \underline{r}_{ij}, \text{ where} \qquad (1)$$

$\underline{Y}_{ij}$ = outcome variable for student_i in school_j,
$\underline{\beta}_{0j}$ = level-1 intercept. $\underline{\beta}_{0j}$ is an adjusted mean for school_j such that
$\underline{\beta}_{0j}=\underline{Y}_{\cdot j}-\underline{\beta}_{1j}(\text{StudentAbility}_{ij})$,
$\underline{\beta}_{1j}$ = Level-1 slope or the expected change in $\underline{Y}_{ij}$ for a unit change in $\text{StudentAbility}_{ij}$, and
$\underline{r}_{ij}$ = residual for student_i in school_j after controlling for student $\text{StudentAbility}_{ij}$.

Level-2

$$\underline{\beta}_{0j} = \gamma_{00} + \gamma_{01}(\text{SchoolAbility}_j) + \underline{\mu}_{0j}, \text{ where} \qquad (2)$$

γ_{00} = Level-2 intercept for $\underline{\beta}_{0j}$,
γ_{01} = Level-2 slope or the expected change in $\underline{\beta}_{0j}$ for a unit change in SchoolAbility_j, and
$\underline{\mu}_{oj}$ = residual for school_j after controlling for SchoolAbility_j.

$$\underline{\beta}_{1j} = \gamma_{10}. \qquad (3)$$

At level-1 we make the assumptions that $E(\underline{r}_{ij}) = 0$, and $Var(\underline{r}_{ij}) = \sigma^2$. At level-2 we assume $E(\underline{\mu}_{0j}) = 0$, $Var(\underline{\mu}_{0j}) = \tau_{00}$.

The above first three equations can be expressed as a single level-1 model by substituting equations 2 and 3 into 1. This yields the reduced form of the HLM model as follows

$$\underline{Y}_{ij} = [\gamma_{00} + \gamma_{01}(\text{SchoolAbility}_j) + \underline{\mu}_{0j}] + [\gamma_{10}](\text{StudentAbility}_{ij}) + \underline{r}_{ij}.$$

Rearranging terms yields

$$\underline{Y}_{ij} = \gamma_{00} + \gamma_{01}\,(\text{SchoolAbility}_j) + \gamma_{10}\,(\text{StudentAbility}_{ij}) + [\underline{\mu}_{0j} + \underline{r}_{ij}]. \qquad (4)$$

The important thing to note about equation 4 is the complicated nature of the error structure [indicated in brackets] which makes it inappropriate for ordinary least squares (OLS).

Identifying Effective Magnet School Programs By Estimating Value-added Effects

One of the main uses of HLM is to provide an index of school program effectiveness. Once the index of school program effectiveness has been estimated then the researcher can rank schools on their effectiveness or use the effectiveness index as a dependent variable to investigate school factors that are related to effectiveness. A very general measure of school program effectiveness can be derived from the simple HLM model provided in equations 1-3.

In the above student level equation, there is one school, $\underline{S}_j$, and one value added effects, $\underline{V}_j$, defined as follows:

$$\underline{S}_j = \gamma_{01}(\text{SchoolAbility}_j) + \underline{\mu}_{0j}\}, \text{ and} \qquad (5)$$

$$\underline{V}_j = \underline{\mu}_{0j}. \qquad (6)$$

$\underline{S}_j$ represents the total effect school programs have on $\underline{Y}_{ij}$, and $\underline{V}_j$ represents the value added by the school program policy to $\underline{Y}_{ij}$. The school effect includes both the effects of contextual variables, (StudentAbility_j), whereas, the value-added effect is the value added after controlling for school context.

Estimates of the school effects and value-added effects may be found by

$$\underline{S}_j = \underline{\beta}_{0j} - \gamma_{00} = \gamma_{01}(\text{StudentAbility}_j) + \underline{\mu}_{0j}\}, \qquad (7)$$

$$\underline{V}_j = \underline{S}_j - \gamma_{01}(\text{StudentAbility}_j) = \underline{\mu}_{0j.} \qquad (8)$$

Both school and value-added effects are calculated after controlling for student background within schools which has already been done in the estimation of $\underline{\beta}_{0j}$ and $\underline{\beta}_{1j}$. A classroom or teacher effectiveness index can be developed by replacing (SchoolAbility_j) with classroom or teacher-level data.

Recent articles on HLM applications are helpful in conceptualizing and explaining this type of analysis to policy makers. For example, Raudenbush and Willms (1995) distinguished between *Type A* and *Type B* school effects. *Type A* effects are often the interest of parents and real estate agents, whereas *Type B* effects are of more interest to education policy makers and evaluators. In a *Type A* effect we consider a school effective when students do well "regardless of whether that school's effectiveness derives from the superb practice of its staff, from its favorable student composition, or from the beneficial influence of the social and economic context of the community in which the school is located. But it would clearly be unfair to reward school staff purely on the basis of their *Type A* effects, given that the staff is only partly responsible for those effects" (p.310). The *Type B* effect is the effect of school practice on student learning unconfounded by school context variables. HLM models are ideally suited to estimate *Type B* effects because they provide an index of school practice variables (curriculum content, instructional practice, and school resources) after factoring out the influence of school context variables and student demographic characteristics. "The *Type B* effect is the effect school officials consider when evaluating the performance of those who work in the schools. A school with an unfavorable context could produce a large *Type B* effect through the effort and talent of its staff. The school would rightly earn the respect of school evaluators even though parents shopping for a large *Type A* effect might not want to choose that school" (p. 310). Equation 7 is the *Type A* school effect and equations 8 is the *Type B* school effect after controlling for between-school context and within-school student effects.

Matrix Notation

The most efficient way of developing the statistical model for HLM and value-added indices is through matrix notation. In the general univariate HLM we have p-1 predictors at level-1

$$\underset{(nj\times 1)}{\underline{\mathbf{Y}}_j} = \underset{(nj\times p)}{\mathbf{X}_j} \underset{(p\times 1)}{\underline{\beta}_j} + \underset{(nj\times 1)}{\underline{\mathbf{r}}_j}, \tag{9}$$

where, $\underline{\mathbf{Y}}_j \sim N(\beta_j, \tau + \sigma^2\mathbf{I}_j)$, $\underline{\beta}_j \sim N(\beta_j, \tau)$, and we assume $\underline{\mathbf{r}}_j \sim N(\mathbf{0}, \sigma^2\mathbf{I}_j)$. $\mathbf{I}_j$ is an $(n_j \times n_j)$ identity matrix.

At level-2 we have q-1 predictors

$$\underset{(p\times 1)}{\underline{\beta}_j} = \underset{(p\times q)}{\mathbf{W}_j} \underset{(q\times 1)}{\gamma} + \underset{(p\times 1)}{\underline{\boldsymbol{\mu}}_j}, \tag{10}$$

where, $\beta_j \sim N(\mathbf{W}_j\gamma, \tau)$ and $\boldsymbol{\mu}_j \sim N(\mathbf{0}, \tau)$. τ is a (p x p) level-2 residual matrix.

The combined model is

$$\underline{\mathbf{Y}}_j = \mathbf{X}_j\mathbf{W}_j\gamma + \mathbf{X}_j\underline{\boldsymbol{\mu}}_j, + \underline{\mathbf{r}}_j, \tag{11}$$

where, $\underline{\mathbf{Y}}_j \sim (\mathbf{X}_j\mathbf{W}_j\gamma, \mathbf{X}_j\tau\mathbf{X}'_j + \sigma^2\mathbf{I}_j)$.

For purposes of estimating value-added effects, we can conceptualize $\mathbf{W}_j$ as consisting of sets of level-2 contextual and treatment variables, $\mathbf{C}_j$ and $\mathbf{T}_j$, with their accompanying regression parameters γ_C and γ_T. Equation 10 can be rewritten as

$$\underline{\beta}_j = \mathbf{C}_j\gamma_C + \mathbf{T}_j\gamma_\mathbf{T} + \underline{\boldsymbol{\mu}}_j. \tag{12}$$

The combined model is

$$\underline{\mathbf{Y}}_j = \mathbf{X}_j\mathbf{C}_j\gamma_C + \mathbf{X}_j\mathbf{T}_j\gamma_\mathbf{T} + \mathbf{X}_j\underline{\boldsymbol{\mu}}_j + \underline{\mathbf{r}}_j. \tag{13}$$

The value added effect $\underline{\mathbf{V}}_j$ is found by

$$\underline{\mathbf{V}}_j = \underline{\beta}_j - \mathbf{C}_j\gamma_C = \mathbf{T}_j\gamma_\mathbf{T} + \underline{\boldsymbol{\mu}}_j. \tag{14}$$

AERA

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030892

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Accountability Evaluation of Magnet School Programs: A Value-Added Model Approach

Author(s): Drs. Eugene P. Adcock and Gary W. Phillips

Corporate Source: Prince George's County Public Schools Upper Marlboro, Maryland

Publication Date: April 25, 2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature:	Printed Name/Position/Title: Dr. Eugene Adcock, Director, REA	
Organization/Address: Prince George's County Public Schools 14201 School Lane, Upper Marlbor, MD 20772	Telephone: 301-952-6240	FAX: 301-952-6199
	E-Mail Address: eadcock@pgcps.org	Date: 4/25/00

(over)

Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

March 2000

Dear AERA Presenter,

Congratulations on being a presenter at AERA. The ERIC Clearinghouse on Assessment and Evaluation would like you to contribute to ERIC by providing us with a written copy of your presentation. Submitting your paper to ERIC ensures a wider audience by making it available to members of the education community who could not attend your session or this year's conference.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed, electronic, and internet versions of *RIE*. The paper will be available **full-text, on demand through the ERIC Document Reproduction Service** and through the microfiche collections housed at libraries around the world.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse and you will be notified if your paper meets ERIC's criteria. Documents are reviewed for contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at **http://ericae.net.**

To disseminate your work through ERIC, you need to sign the reproduction release form on the back of this letter and include it with **two** copies of your paper. You can drop of the copies of your paper and reproduction release form at the ERIC booth (223) or mail to our attention at the address below. **If you have not submitted your 1999 Conference paper please send today or drop it off at the booth with a Reproduction Release Form.** Please feel free to copy the form for future or additional submissions.

Mail to: AERA 2000/ERIC Acquisitions
The University of Maryland
1129 Shriver Lab
College Park, MD 20742

Sincerely,

Lawrence M. Rudner

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

ERIC/AE is a project of the Department of Measurement, Statistics and Evaluation at the College of Education, University of Maryland.